COMPUTER METHOD AND APPARATUS FOR UNIFORM
REPRESENTATION OF GENOME SEQUENCES

ABSTRACT OF THE DISCLOSURE

A method and apparatus transforms typically differing length text string

5   representations (i.e., sequences) of biological fragments into uniform length

representations. A comparison database stores a predefined number of known

biological sequences. A comparison routine compares and scores a subject sequence

against each known sequence in the database. Each individual score (one for each

known sequence in the database) serves as a vector element forming a fixed length

10   vector representation of the subject sequence. Vector length equals the predefined

number of known biological sequences in the database. Scoring is a probability or an

occurrence count of the known biological sequence in the subject sequence.